

(19)



Europäisches Patentamt
European Patent Office
Office européen des brevets



(11)

EP 1 400 953 A1

(12)

EUROPEAN PATENT APPLICATION

(43) Date of publication:
24.03.2004 Bulletin 2004/13

(51) Int Cl.7: G10L 15/28

(21) Application number: 02405798.6

(22) Date of filing: 12.09.2002

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR
IE IT LI LU MC NL PT SE SK TR
Designated Extension States:
AL LT LV MK RO SI

(71) Applicant: Swisscom AG
3050 Bern (CH)

(72) Inventor: Van Kommer, Robert
1752 Villars-sur-Glâne (CH)

(74) Representative: Saam, Christophe
Patents & Technology Surveys SA
Terreaux 7
Case postale 2848
2001 Neuchâtel (CH)

(54) Method for building speech and/or language recognition models

(57) A method for building the speech and/or language models (41, 42) which are initially loaded in a multimodal handheld device (4), such as a PDA, a mobile phone or a toy. The speech and language models are used for speech recognition. The method comprises the following steps:

said speaker selects (406) on a visual user interface

of said multimodal handheld device one or several speaker clusters to which said speaker belongs, said multimodal handheld device (4) sends (408) said selected clusters to a remote server (1), cluster-dependent speech and/or language models (41, 42) are downloaded (412) from said remote server (1) to said multimodal handheld device (4).

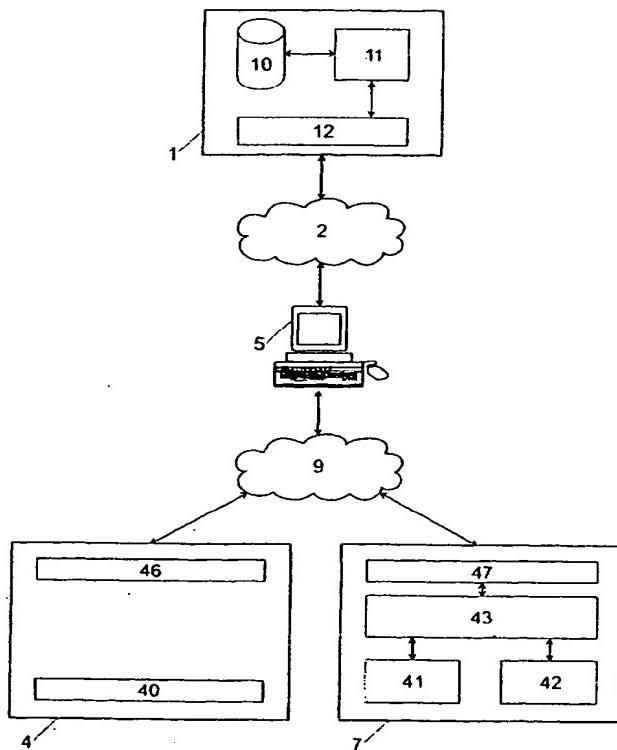


Fig. 4

Description

[0001] The present invention concerns a method for building speech and/or language recognition models. In particular, the present invention concerns a method adapted to multimodal handheld devices such as PDAs (personal digital assistants), mobile phones, PDAs combined with mobile phones, and toys.

[0002] PDA have established themselves as a convenient replacement of paper diaries. Many PDAs also offer a wide range of software applications and services for people on the move. Modern PDAs include a GPRS (General Packet Radio Service), UMTS (Universal Mobile Telecommunication System) or Wireless LAN IEEE 802.11 b interface and are therefore capable of being always connected to a telecommunication network ("always on").

[0003] For a broad market acceptance, PDAs should be kept small and light-weight. This implies severe restrictions on the input means that can be integrated. The present solutions include for example soft keyboards or keypads or pen-based handwriting recognition systems. While pen-based selection in a visual menu is convenient, getting longer texts into a PDA is hardly practicable.

[0004] Speech recognition based systems have already been used with success as a complementary input mode to enhance and complement the pen-based interface. Speech is very natural and easy to use; furthermore, the text acquisition speed is much higher than with most soft keyboards or pen-based input systems. For cars drivers, speech recognition provides a hands-free and eyes-free interface.

[0005] Experiences with existing systems show that speech input is an ideal complement for pen-based handheld devices; people tend to prefer speech for entering data and pen or buttons for corrections and pointing.

[0006] Most existing speech recognition modules in handheld devices use neuronal networks or Hidden Markov Models (HMMs) with speaker-independent speech recognition models. For each speech element in a dictionary of words to be recognised, a speech model is stored in the handheld device. Predefined speech models are stored in the system during the installation of the speech recognition module. Speech elements can include for example words, subwords, phonemes, triphones, polyphones or sentences.

[0007] Updates or additions of new speech elements in current speaker-independent speech recognition systems are impossible or at least cumbersome. Moreover, the speech recognition performance of speaker-independent speech recognition systems remains usually poor and the vocabulary of words accessible by voice is limited.

[0008] Some handheld devices achieve better performance by using speaker-dependent speech models, which are usually learned during an enrolment and/or training session. During this phase, the speaker has to

teach his handheld device how he pronounces words and commands. New words, for example new commands or new contact names or personal contacts in an agenda, can be inputted and must be trained. The error rate of speaker-dependent speech recognition systems is usually much lower than for speaker-independent speech recognition systems. However, many users view the time spent for the enrolment session as a burden, which is a serious barrier to broad market acceptance.

[0009] Adaptative speech recognition methods have been proposed for other applications than for recognition purposes in handheld devices. Adaptative methods preclude the need of an enrolment or training session and still reach very high recognition rates. In those systems, initially available speaker-independent speech models are progressively adapted to the specific voice and language of the speaker, using speech elements, words and sentences spoken by the speaker during normal working sessions. The adaptation may be guided - the speaker has to correct the recognition system each time a word has been wrongly recognized - or unguided - the recognition system corrects itself and improves the speech models automatically.

[0010] Adaptative speech recognition systems are widely used, for example in speech recognition software for personal desktop computers. However, for different reasons, they are only poorly adapted to speech recognition in handheld devices. It can be shown that the performance of adaptative speech recognition methods in handheld devices, in particular the initial recognition rate before adaptation, remains poor.

[0011] One first reason is that storage capacity in handheld devices, for example in PDAs, mobile phones or toys, is scarce in comparison with desktop computers. Consequently, the vocabulary of trained words or speech elements for which a model is initially available should be kept small. Speech elements that do not belong to the initial dictionary are not recognised; users of a new, untrained system still have to create models for all the speech elements absent from the initial set.

[0012] Besides, as speech models built from utterances of large groups of speakers tend to necessitate more storage space, the speech models initially loaded in handheld devices are often created from smaller data sets, which also decrease the initial recognition performances.

[0013] As a consequence, the recognition rate of a new, not yet adapted speech recognition module in a handheld device is often so poor that many users who only give it a few tries are discouraged from using it.

[0014] One aim of the invention is therefore to build more easily, more quickly and with less user intervention improved, pre-adapted speech and/or language models for speech recognition in a handheld device.

[0015] Another aim of the invention is to improve the recognition rate of speech recognition modules in handheld devices, more particularly before the adaptation.

[0016] Another aim of the invention is to adapt the

speech and/or language models available in a handheld device without necessitating an enrolment session.

[0017] In accordance with one embodiment of the present invention, those aims are achieved with a method for constructing speech and/or language recognition models during a speaker enrolment and/or training phase for a multimodal handheld device, wherein the speaker first selects on a visual user interface of said device one or several speaker clusters to which said speaker belongs. The handheld device sends said selected clusters to a remote server. The remote server computes or retrieves cluster-dependent speech and/or language models for speech elements in a cluster-dependent dictionary that are downloaded to said device.

[0018] One advantage is that the speech and/or language models initially loaded in the handheld device, and the dictionary of speech elements for which models have been built, are already pre-adapted to the characteristics of the user and are therefore close to the user specifications. The initial recognition rate that can be reached with the models downloaded in the handheld device prior to the first use of its speech recognition module is therefore higher than in known enrolment-less systems in which speaker-independent models are initially used.

[0019] Furthermore, as the speech and/or language models are built from voice data gathered from speakers belonging to the same clusters as the user of the handheld device, high recognition rates can eventually be achieved even with small speech and/or language models built from small sets of speakers.

[0020] Different types of speaker clusters can be defined. Best results will be reached if the speakers in each cluster have similar speech characteristics or models. Preferably, one will use clusters which the speakers will have no difficulties in selecting for themselves. Examples of suitable clusters may include sex, age, mother tongue, geographic location or origin, weight, education level and/or main professional activity of the speaker. Obviously, one speaker may chose to belong to different independent clusters in order to further improve the initial recognition rate before adaptation. For example, one speaker may define himself as a man, aged between 30 and 40, having English as a mother tongue, born in Geneva, weighing between 70 and 80 KGs, working as a marketing assistant.

[0021] If a speech sample has already been recorded, the system can help the user to select the right cluster, or check if the selection of categories is appropriate.

[0022] The invention will be better understood with the help of the description of a specific embodiment illustrated by the figures in which:

Fig. 1 shows a diagram of a first embodiment of a telecommunication system according to the invention.

Fig. 2 shows a second embodiment of a telecom-

munication system according to the invention.

Fig. 3 shows a third embodiment of a telecommunication system according to the invention.

Fig. 4 shows a fourth embodiment of a telecommunication system according to the invention.

Fig. 5 shows a fifth embodiment of a telecommunication system according to the invention.

Fig. 6 is a flowchart of the method of the invention.

Fig. 7 illustrates one visual interface that can be used for implementing the method of the invention with a handheld device.

[0023] Fig. 1 shows a diagram of a telecommunication system according to the invention. It includes a remote server 1 operated by a service provider, for example by a mobile network operator, by an internet service provider or by a software solution provider and serving a plurality of users (or subscribers) using multimodal handheld devices 4. The remote server includes a database 10 for previously stored speech data units, i.e. a collection of audio data corresponding to speech elements uttered each by preferably several hundreds or thousands of persons. According to the invention, clusters (as later defined) to which the speaker of each audio data unit belongs are linked in the database 10 with the corresponding speech data units.

[0024] A speech and/or language-computing module 11 is provided for computing speech and/or language models corresponding to possible combinations of clusters in the database 10. The module 11 can send a query to the database 10 for retrieving all the speech data units corresponding to a selected combination of speaker clusters. For example, the module 11 can retrieve all speech data units recorded by female speakers less than 20 years old. Speech and language models are computed from those speech data units that can be used by an independent or adaptative speech recognition module.

[0025] In the illustrated embodiment, the server 1 further comprises an Internet server, for example an http or https server for connecting it to the Internet 2. Internet users can download speech and/or language models computed by the module 11 by connecting to a corresponding page on the https server 12, as will be described later. Speech samples can be uploaded for automatically suggesting clusters or verifying selected clusters.

[0026] The system of the invention comprises a plurality of multimodal handheld devices 4; for the sake of simplicity, only some components of a single device 4 are illustrated on Figure 1. In this description and in the claims, multimodal means that the handheld device 4 comprises at least two different input means 40 for in-

putting commands or data, one of those means comprising a microphone and a preferably adaptative speech recognition module 43 for inputting spoken commands and texts. Depending on the device, the other input means 40 can comprise, for example, a pen-based interface with a handwriting recognition module, a keypad, a soft keyboard, buttons, joysticks, etc or any combination of those means.

[0027] In this first embodiment, the handheld device 4 comprises a radio interface 44 for connecting it to a mobile data communication network 3. The network 3 can for example be a wireless LAN, for example a IEEE 802.11b LAN, a GSM (Global System Mobile), GPRS (General Packet Radio Service), HSCSD (High Speed Circuit Switched Data), Edge or UMTS (Universal Mobile Telecommunication System) network; different data communication layers, for example WAP (Wireless Application Protocol) or I-Mode, can be established over this network. The device 4 can connect itself to the Internet 2 through the mobile network 3 and through an Internet Service Provider 20. The interface module 44 can be internal, for example if the handheld device consists of a mobile phone or is connected to the handheld device 4 over an IrDA, PCCard, Bluetooth, WLAN, serial or direct connection.

[0028] The speech recognition module 43 preferably comprises a software application executed by the general-purpose microprocessor (not shown) of the handheld device for providing speech recognition services to other applications. This module 43 is preferably written in an object-oriented language, for example C++, IDL or JAVA (registered mark of SUN Microsystems), preferably using distributed components, for example using CORBA or a web services description language, but can also comprise modules written in an iterative language. It preferably uses adaptative Hidden Markov Models (HMMs) or adaptative neuronal networks for recognising speech elements spoken by the user of the handheld device and for converting them into commands and/or text input in various applications. The module 43 uses adaptative speech models 41 and adaptative language models 42. Speech models define the way each speech element is pronounced by the speaker. Language models define typical expressions and syntaxes often used by the speaker. The speech recognition module 43 uses the models 41, 42 to recognise speech elements, including words, commands and sentences spoken by the user; furthermore, it continuously adapts those models, in a guided or preferably unguided way, in order to improve the recognition rate reached by the user of the handheld device.

[0029] The one skilled in the art will understand that, if a plurality of users share the same device 4, a corresponding plurality of speech and language models 41, 42 can be defined and stored in the same handheld device 4.

[0030] Figure 2 illustrates a second embodiment of the system of the invention. Features which are identical

to those of Figure 1 share the same reference numbers and will not be described again. In this embodiment, the illustrated handheld device 4 lacks the module for directly connecting it to a cellular data network; instead, it can be connected over an interface 45 to a terminal 5, for example a personal computer, which is itself connected to the Internet 2. In a preferred embodiment, the connection between the handheld device 4 and the terminal 5 is achieved by inserting the device 4 on a cradle (not shown) connected to the terminal 5 over an USB link, for example. Other ways of connecting a handheld device, for example a PDA or a mobile phone, to a personal computer can easily be conceived by the one skilled in the art. As a non-exhaustive list of examples, we will cite

5 a direct USB link, a wireless short-range interface (for example an infrared or Bluetooth interface), a serial interface, a wireless LAN (WLAN) according to 802.11b or HomeRF, etc.

[0031] Figure 3 illustrates a third embodiment of the 10 system of the invention. Features which are identical to those of Figure 1 and/or 2 share the same reference numbers and will not be described again. In this embodiment, the speech and language models are retrieved from a server 5, for example a personal computer, connected with the multimodal handheld device 4 over a temporary, for example serial, or permanent, for example in a local LAN, connection. In a preferred embodiment, the speech and/or language models 10 are computed by the server 5 using a software application 11

15 stored on a storage medium 50, for example a floppy or an optical storage medium. The software application uses a database 10 for speech data units that can be stored together with the application on the storage medium 50 or, as an alternative, retrieved over the Internet. 20 If the database 10 is locally available, this embodiment is particularly suitable for users who lack a fast Internet connection. Alternatively, it would also be possible to retrieve the parameters for speech recognition corresponding to the selected clusters from a storage medium 25 directly connected to the multimodal handheld device.

[0032] The performance, including the processing power and the storage space, of many handheld devices is usually smaller than that delivered by personal computers or servers in a LAN environment. In some cases, those performances may be insufficient for efficiently implementing a software-based speech recognition algorithm. Figure 4 illustrates a fourth embodiment of the invention that addresses this problem. Features that are identical to those of the preceding figures share the same reference numbers and will not be described again. In this embodiment, the multimodal handheld device 4 is permanently connected over a wireless local area network 9 to a local server 7. The speech recognition module 43 and the speech and/or language models 41, 42 initially downloaded from a remote server 1 are stored and run by the server 43. In this embodiment, speech recognition services requested by the multimodal handheld device 4 are provided by the local server 7.

dal handheld device 4 are delivered by the locally connected server 7. Distributed programming protocols, including web services, may be used for programming this embodiment.

[0033] Figure 5 illustrates a fifth embodiment of the system of the invention. Features that are identical to those of the preceding figures share the same reference numbers and will not be described again. In this embodiment, a complete speech recognition system is stored and run by a local server connected to the multimodal handheld device 4 over a wireless LAN 4, as in the fourth embodiment. However, a simpler speech recognition module 43', and/or a smaller set of speech models 41', and/or a smaller set of language models 42', is stored and can be run by the multimodal handheld device 4. In this embodiment, the speech recognition system 41', 42' and 43' of the handheld device is used when there is no access to the server 7, or when this system is reliable enough for the current application, and the more powerful speech recognition system 41, 42, 43 of the server 7 is used when a more reliable speech recognition is needed.

[0034] The one skilled in the art will recognize that it is also possible to have the speech recognition module 43 and/or the models 41, 42 in a server accessible over a public network, for example over the Internet, when a fast permanent connection is available.

[0035] Figure 6 is a flowchart illustrating a preferred embodiment of the method of the invention. In a first step 400, the speech recognition application module is launched in the handheld device. This may occur for example because the user starts an application on his handheld requiring speech recognition, or because the user has pressed a special speech button 40 (Figure 7) for turning on the microphone 46, recording the voice and converting it to text.

[0036] During step 402, the speech recognition application module checks if speech and language models are available. This should always be the case, except before the first use of the speech recognition module.

[0037] If the speech and language models 41, 42 are available, the process continues with step 414 during which the recorded speech is converted to text or commands using known speech analysing methods and the available models 41, 42. The converted text or command is sent to the requiring application.

[0038] During step 416, which can be combined with step 414, the speech and language models 41, 42 are adapted to the speech and language of the user of the handheld device 4. Adaptation may be performed using known speech models adaptation methods and may be guided (i.e. the user acknowledges or corrects the text converted by the speech conversion module) or unguided (the module detects and automatically adapts the speech and/or language models in order to improve the recognition confidence rate). When all the spoken speech elements have been recognised, the speech recognition application module ends at step 418.

[0039] If the module detects during step 402 that speech and/or language models are not yet available for the user currently logged in, it continues with the speaker clusters selection step 404. This step may include the execution of a dedicated speaker clusters selection routine in the handheld device 4, or the connection with the handheld device 4 to a speaker clusters web, WAP or I-mode page on https server 12. In another embodiment, the speaker clusters selection step is

5 performed with the terminal 5, for example with a dedicated application in the terminal 5 or by accessing a dedicated web page on https server 12 with terminal 5.

[0040] During step 406, the speaker actually selects speaker clusters to which he believes he belongs. This

10 step is illustrated on Figure 7. Various types of speaker clusters are presented to the speaker on the display 47 of the handheld device 4. In this example, the available clusters include sex, age, mother tongue, geographical origin, weight, education level, main professional activity. The speaker may also select during this step the language of the desired speech models (not shown). The one skilled in the art will have no difficulties in imagining other types of predefined clusters for subdividing the speakers into different classes sharing common speech

15 or language properties. For each cluster type, the user can enter or select in a menu on the visual interface of his multimodal handheld device values corresponding to his cluster. In a preferred embodiment, clusters are selected with a pen-based interface directly on the display 47 of the handheld. Other selection means may be available. Clusters can be suggested by the server 1 if user's speech samples have been previously uploaded.

[0041] During step 408, the selected clusters are sent to the remote server 1. If the values have been selected on a web page of the server 12, this step simply implies to send a fulfilled web form to the server 12. If the values have been selected with a dedicated routine in the handheld device 4, this step may include sending a message, for example an email or SOAP message, to or establishing a connection with the remote server 1.

[0042] When it receives the list of clusters selected by the user of a handheld device, the remote server 1 sends a query to the speech database 10 for retrieving the speech data of the set of speakers corresponding to the selected combination of clusters. It then computes speech models and/or language models corresponding to the selected set of speakers. In order to compute efficient speech models, the set preferably includes at least several dozens, if possible several hundreds, different speakers. If storage space is an issue, the dictionary of speech elements for which models are built also depends on the selected combination of speaker clusters. For instance, it would be possible to download a dictionary of words including a different professional jargon for each professional categories selected. It may also compute or retrieve different probabilities of occurrence associated with each speech element, for example with each word, and/or with each sequences of

speech elements, depending on the selected cluster.

[0043] If the speech recognition module of the handheld device uses specific grammars, for example for retrieving dates or other formatted data units, the set of downloaded grammars may also depend on the selected clusters. In an embodiment, it is also possible to adapt the parameters needed for the speech features extraction to each cluster, for example to extract different features from a woman's speech sample than from a man's in order to take into account different psychoacoustic parameters.

[0044] If speech samples from the user are available, the server 1 may check if the clusters selected by the user are appropriate.

[0045] The handheld device 4 waits during step 410 for an answer from the remote server 1. A time limit may be set after which the selected clusters will be sent again or the process will be stopped.

[0046] If a positive answer is received, the speech recognition module 43 in the handheld device 4 downloads during step 412 the parameters for the speech recognition corresponding to the combination of clusters selected by the speaker. Those parameters preferably include speech and/or language models made of one or several data files in a format depending on the speech recognition module used by the handheld device 4.

[0047] As the speakers selected for constructing the downloaded speech and language models belong to the same clusters as the user of the multimodal handheld device 4, the downloaded initial set already allows an acceptable recognition rate of speech elements spoken by this user, even if the number of speakers used for building this speech model is limited.

[0048] A speech and/or language models check can be performed to verify the downloaded models (not shown). The user has to speak a few sentences, for example sentences displayed on the visual interface of his handheld device 4, which the speech recognition module tries to recognize with the downloaded models. If the recognition rate is too low, or if the recognition confidence level is below a predefined threshold, the models are rejected and the speech recognition module prompts the user to select another combination of clusters. The system may also suggest better or more appropriate clusters, or use a standard speech model.

[0049] In another embodiment (not shown), the check of the speech and language models built from the selected combination of speaker clusters is performed in the remote server 1 prior the downloading of said speech and language models. For this, speech data recorded from the speaker 1 have to be sent to the remote server 1.

[0050] Once the speech and language models have been accepted, the speech recognition application module goes on with the already described steps 414 to 418 for recognizing speech elements and commands uttered by the user.

[0051] In all the above-described embodiments, the

remote server 1 is made of a central server accessible by a plurality of users over the Internet. The one skilled in the art will understand that the same functions can also be performed by a server, for example by the personal desktop computer 5, in the vicinity of the user. It would be possible for example to store the speech data

10 and the speech and/or language computing module 11 on an optical storage disk sold with the speech recognition application module 44 and loaded in the computer 5 for providing the same functionality. This allows an easier downloading of speech and language models for users who do not have an Internet connection, but has the drawback that late adaptation of the contents of the database 10, for example in order to include new speech elements or new commands corresponding for example to new software applications available, are much harder to distribute to a large community of users.

[0052] Furthermore, the one skilled in the art will understand that different portions of the speech and language models can be downloaded from different remote servers 1. This allows for example a software provider to distribute over his web site pre-adapted speech and language models corresponding to the command words needed for running his software. It may also be possible 25 to have different servers for different languages of the speech models.

Claims

30 1. A method for building speech and/or language models (41, 42) used for recognition of speech spoken into a multimodal handheld device (4), said method comprising the following steps:

35 a speaker selects (406) on a visual user interface (47) of said multimodal handheld device (4) one or several speaker clusters to which said speaker belongs, cluster-dependent parameters (41, 42) for the speech recognition are computed and used (412) for said speech recognition.

40 2. The method of claim 1, wherein said parameters for the speech recognition include cluster-dependent speech and/or language models (41, 42).

45 3. The method of one of the claims 1 or 2, wherein said parameters for the speech recognition are downloaded from a remote server (1).

50 4. The method of claim 3, wherein said parameters for the speech recognition are downloaded into said device (4) over one of the following networks: GSM, GPRS, HSCSD, UMTS, WLAN.

55 5. The method of one of the claims 1 or 2, wherein said parameters for the speech recognition (41, 42) are

- retrieved from a storage medium (50) non-permanently connected to said multimodal handheld device.
6. The method of one of the claims 1 or 2, wherein said parameters for the speech recognition (41, 42) are stored in a server (7) permanently accessible from said multimodal handheld device.
7. The method of claim 1, wherein said speaker categories comprise at least two categories chosen among the following:
- sex
age
mother tongue
geographic location or origin
weight
education level
main professional activity,
- said speech and/or language models being computed from speech data of speakers corresponding to the selected combination of categories.
8. The method of one of the claims 1 to 7, further comprising the following steps:
- said speaker speaks in order to check if the selected cluster-dependent parameters for the speech recognition (41, 42) are indeed appropriate.
9. The method of step 8, said sentences being sent to said remote server (1) for performing a remote check prior to said downloading step.
10. The method of one of the claims 1 to 7, further comprising a step of using available speech data from the speaker to suggest some clusters to said speaker.
11. The method of one of the claims 1 to 10, further comprising a step of adapting (416) said parameters for the speech recognition (41, 42) to said speaker.
12. The method of one of the claims 2 to 11, wherein the list of speech elements for which speech and/or language models (41, 42) are made available depends on said selected clusters.
13. The method of one of the claims 1 to 12, wherein said remote server (1) selects pre-recorded speech data units (10) corresponding to speakers belonging to the selected combination of clusters in order to compute on-the-fly said cluster dependent speech and/or language models (41, 42).
14. The method of one of the claims 1 to 13, said clusters being selected prior to the first use of the speech recognition module (44).
- 5 15. The method of one of the claims 1 to 14, said clusters being selected on a web page in said multimodal handheld device (4).
- 10 16. The method of one of the claims 1 to 15, said clusters being selected with a dedicated speaker clusters selection routine in said multimodal handheld device (4).
- 15 17. The method of one of the claims 1 to 16, said clusters being selected with a pen-based interface on said multimodal handheld device (4).
- 20 18. A server, comprising:
- means for connecting it to a public telecommunication network (2),
a database (10) containing speech data units, each speech data unit being linked to several clusters to which the speaker of said data unit belongs,
computing means (11) for computing parameters for the speech recognition from a subset of speech data records in said database corresponding to various combinations of clusters.
- 25 30 19. The server of claim 18, further comprising an Internet server (12) over which said computed parameters for the speech recognition can be downloaded.
- 35 40 20. The server of claim 19, wherein said Internet server includes at least one page with which remote users can define a combination of clusters for which parameters for the speech recognition have to be computed.
- 45 50 55

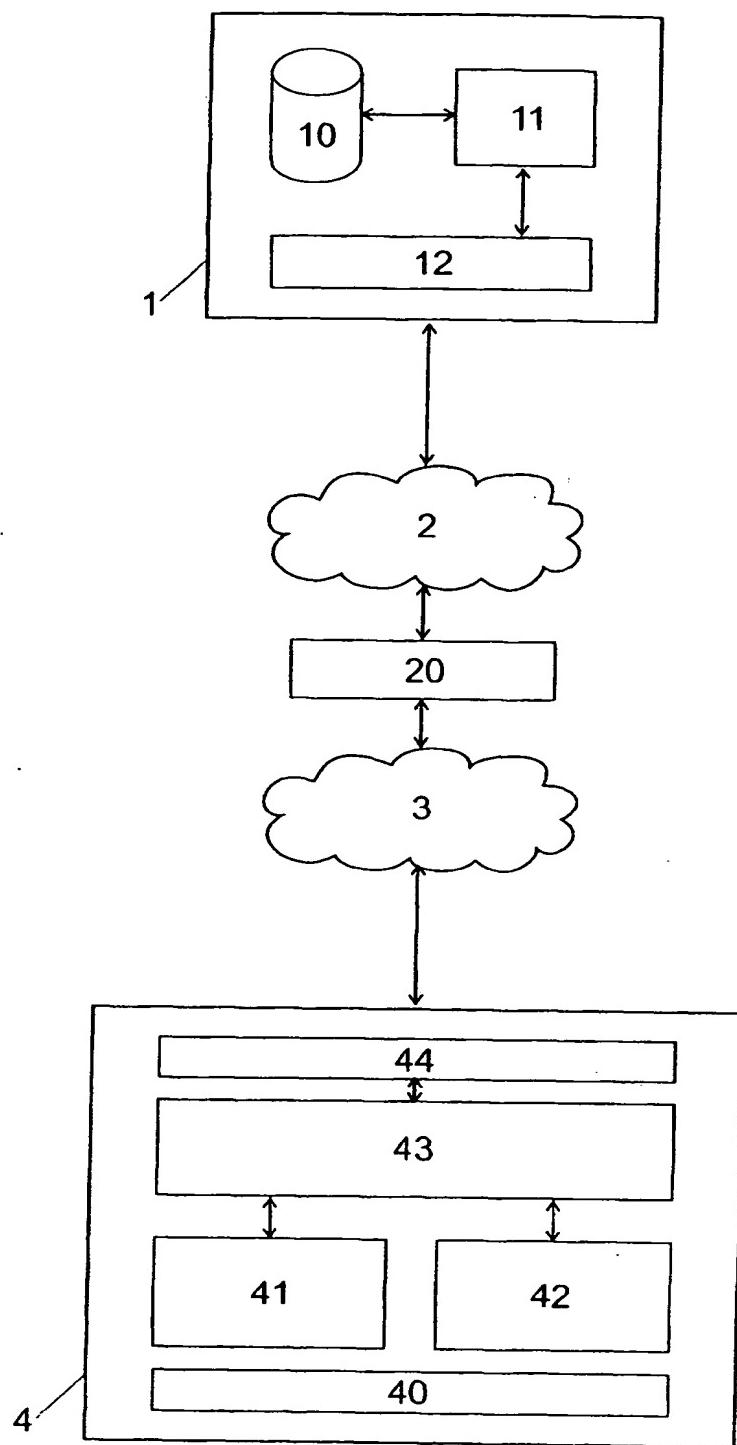


Fig. 1

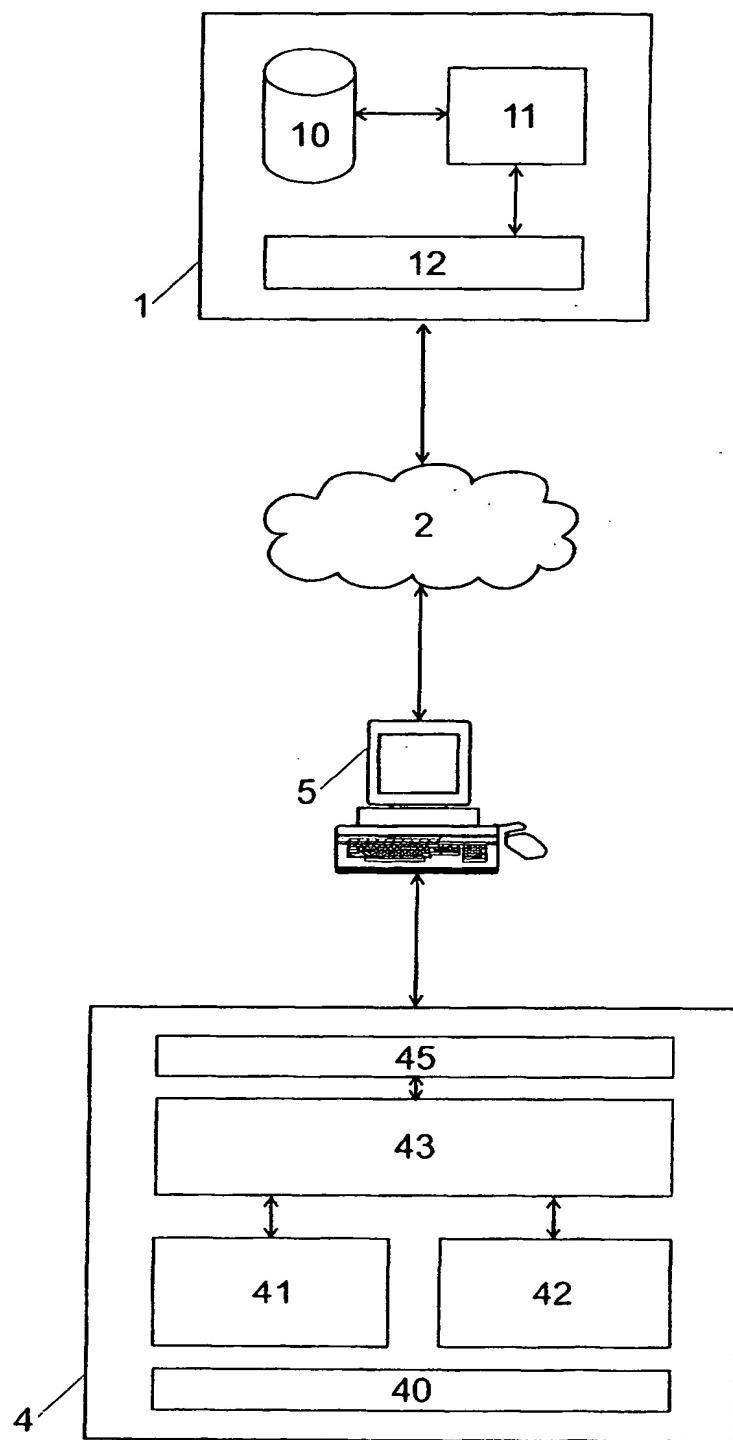


Fig. 2

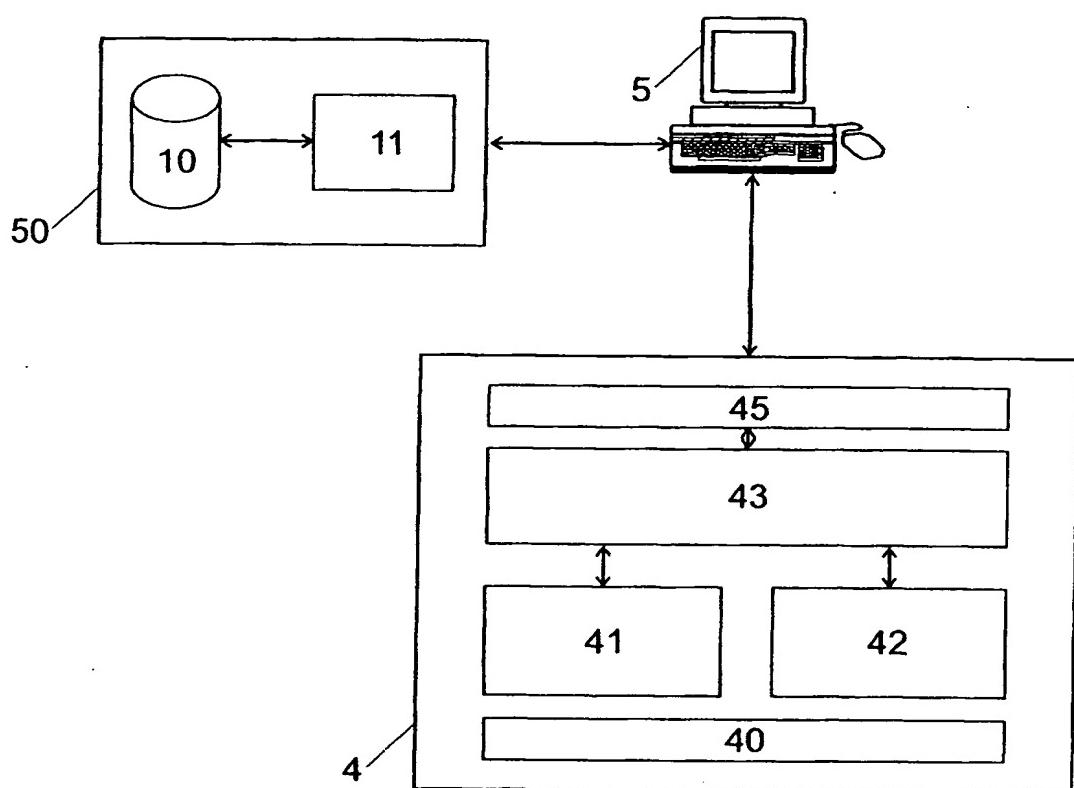


Fig.3

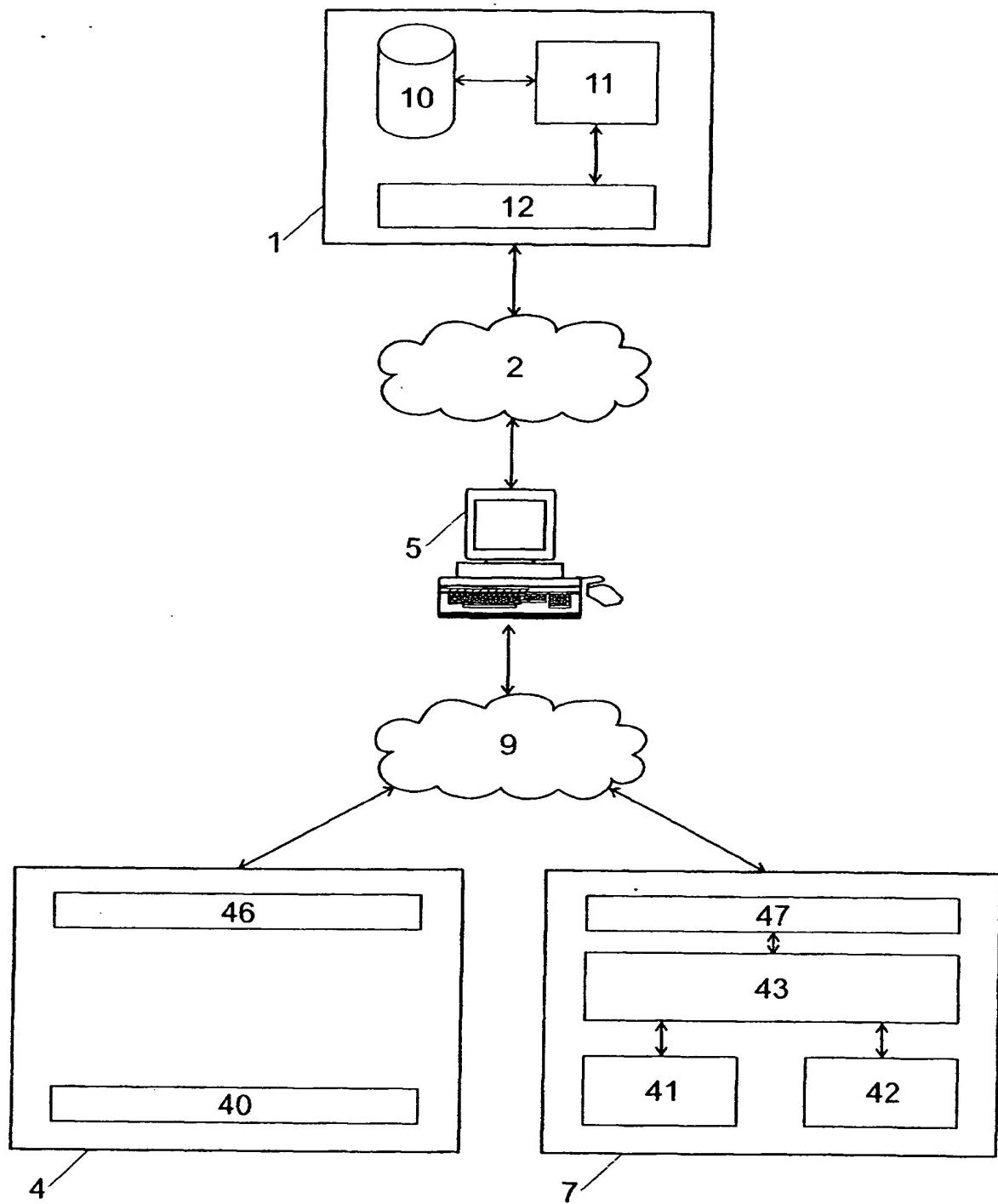


Fig. 4

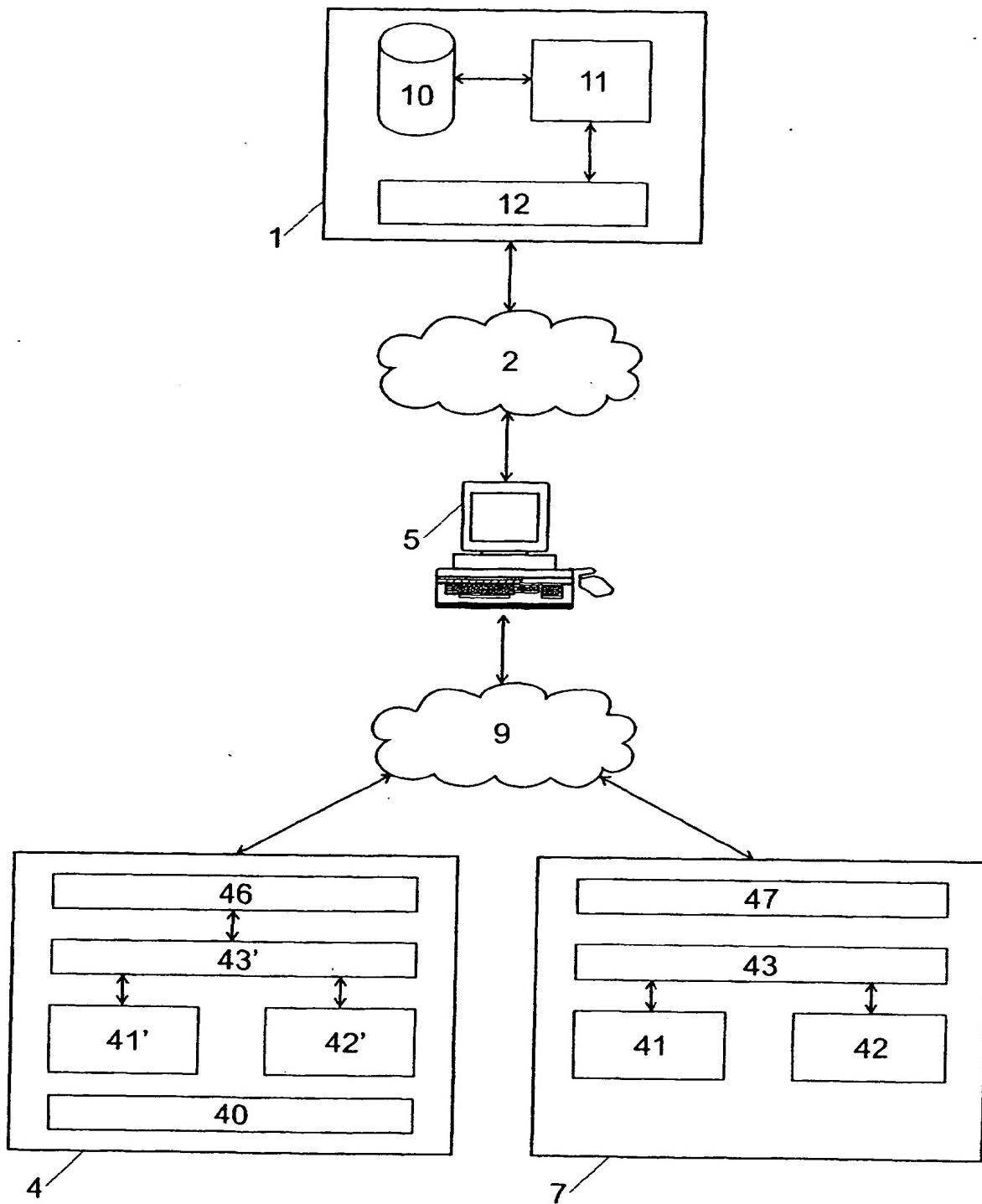


Fig. 5

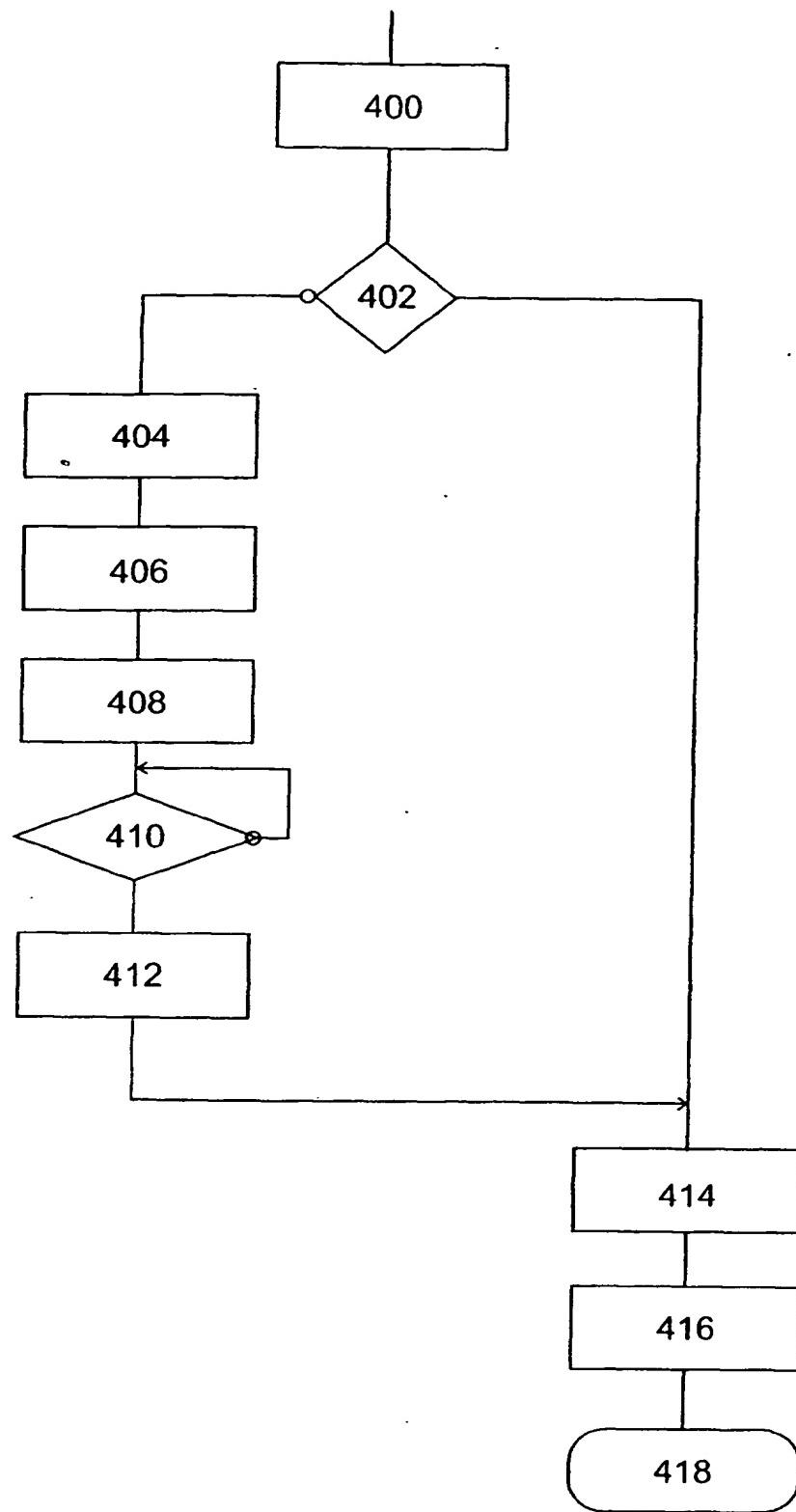


Fig. 6

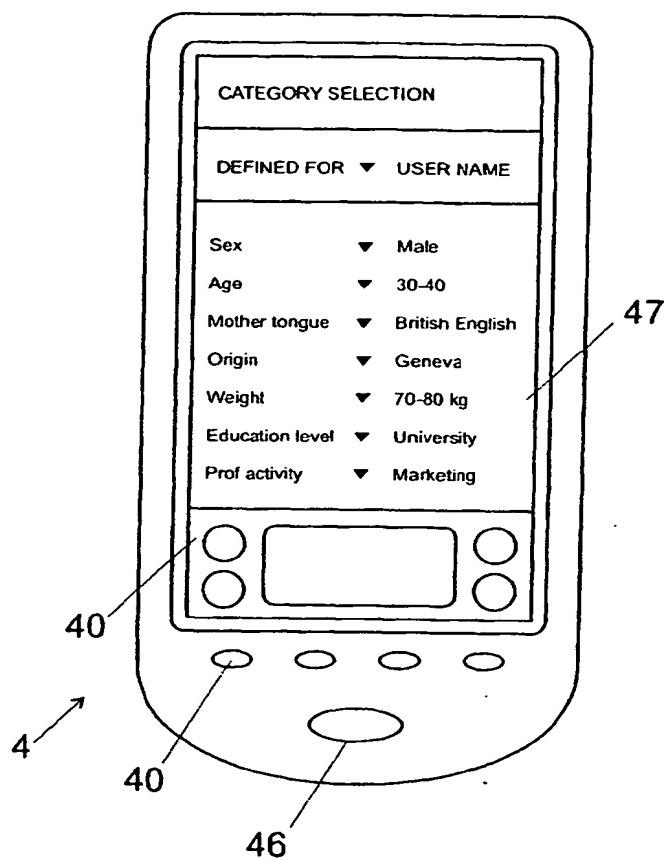


Fig. 7



European Patent
Office

EUROPEAN SEARCH REPORT

Application Number
EP 02 40 5798

DOCUMENTS CONSIDERED TO BE RELEVANT			CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	
X	US 6 393 403 B1 (MAJANIEMI MARKKU) 21 May 2002 (2002-05-21) * abstract; figures 1-3 * * column 2, line 53 - column 3, line 12 * * column 4, line 46 - column 5, line 22 * ----	1-7,12, 14,16, 18,19	G10L15/28
A	US 5 905 773 A (WONG CHI) 18 May 1999 (1999-05-18) * abstract; figure 5 * * column 2, line 44 - line 57 * ----	1,18	
A	US 5 960 063 A (KUROIWA SHINGO ET AL) 28 September 1999 (1999-09-28) * column 2, line 26 - column 3, line 20; figure 1 * -----	1,18	
			TECHNICAL FIELDS SEARCHED (Int.Cl.7)
			G10L
<p>The present search report has been drawn up for all claims</p>			
Place of search	Date of completion of the search	Examiner	
THE HAGUE	20 January 2003	Wanzeele, R	
CATEGORY OF CITED DOCUMENTS			
X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document			
T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document			

**ANNEX TO THE EUROPEAN SEARCH REPORT
ON EUROPEAN PATENT APPLICATION NO.**

EP 02 40 5798

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

20-01-2003

Patent document cited in search report		Publication date		Patent family member(s)		Publication date
US 6393403	B1	21-05-2002	NONE			
US 5905773	A	18-05-1999	CA WO EP	2250050 A1 9737481 A1 0890249 A1	09-10-1997 09-10-1997 13-01-1999	
US 5960063	A	28-09-1999	JP GB	10063293 A 2316575 A ,B	06-03-1998 25-02-1998	